

ADMN: A Layer-Wise Adaptive Multimodal Network for Dynamic Input Noise and Compute Resources

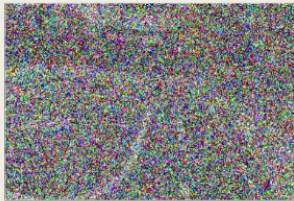
Jason Wu*, Yuyang Yuan*, Kang Yang*, Lance Kaplan^, Mani
Srivastava*

*UCLA

^DEVCOM Army Research Laboratory

Motivation – Dynamic World State

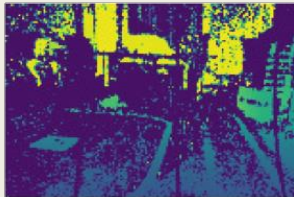
Noisy Image



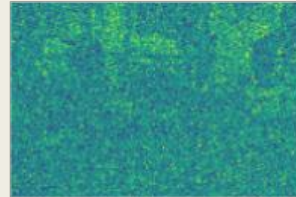
Clean Image



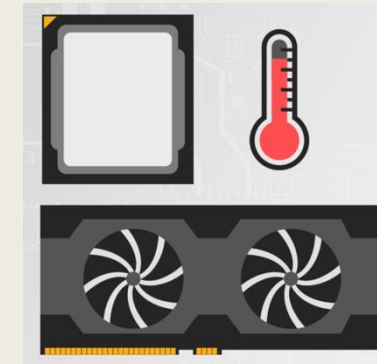
Clean Depth



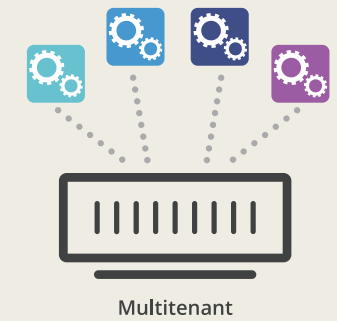
Noisy Depth



Dynamic Modality Quality of
Information (QoI)

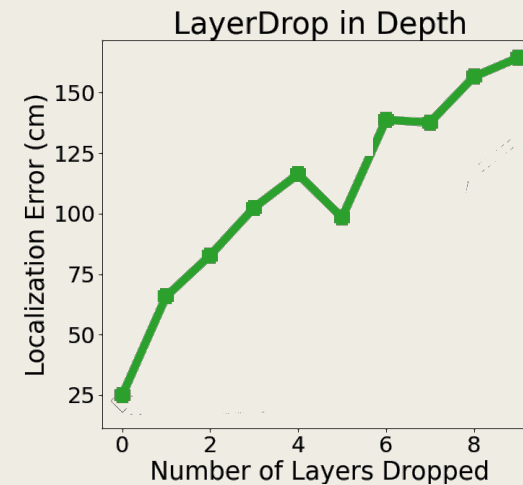
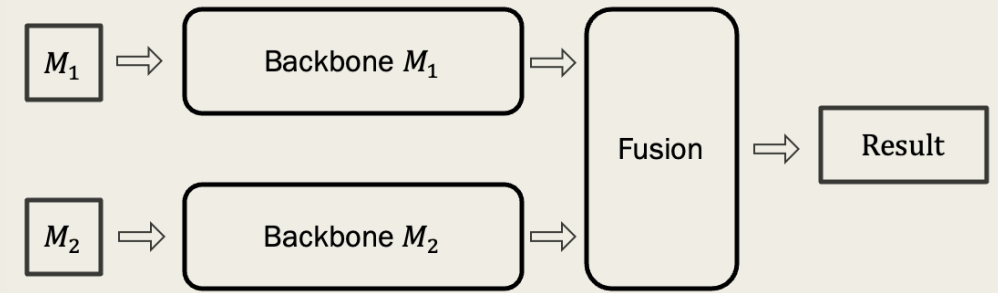


Fluctuating Computational
Resource Availability

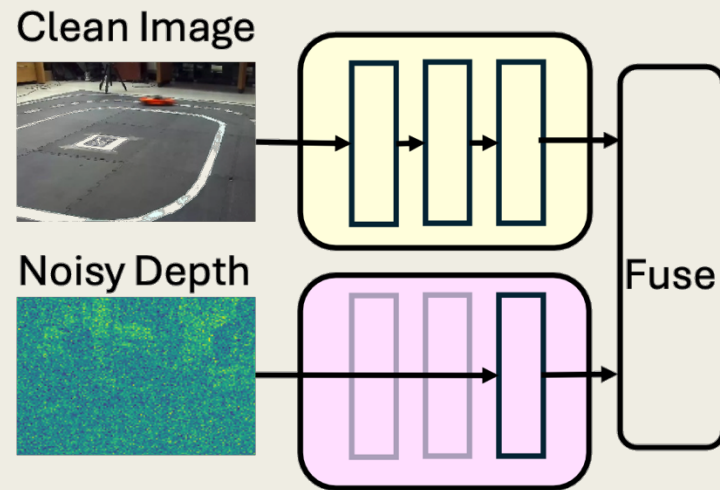


Challenges

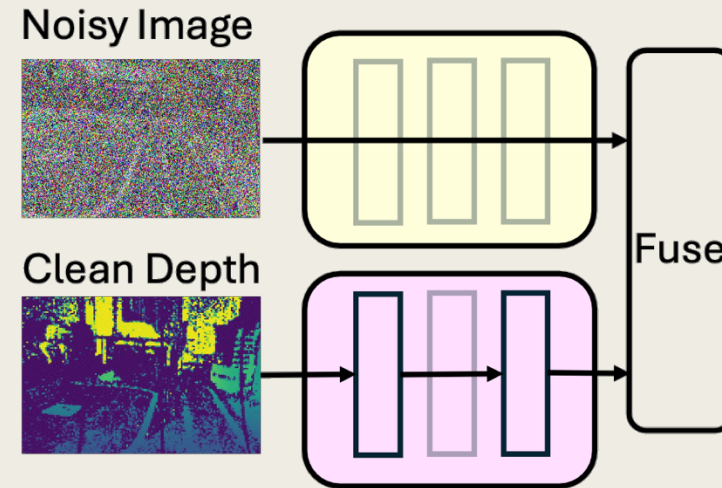
- Existing multimodal networks are largely static
- Modalities are allocated the **same resources** regardless of world state
- Cannot accommodate variable computational resource availability



Proposed Solution

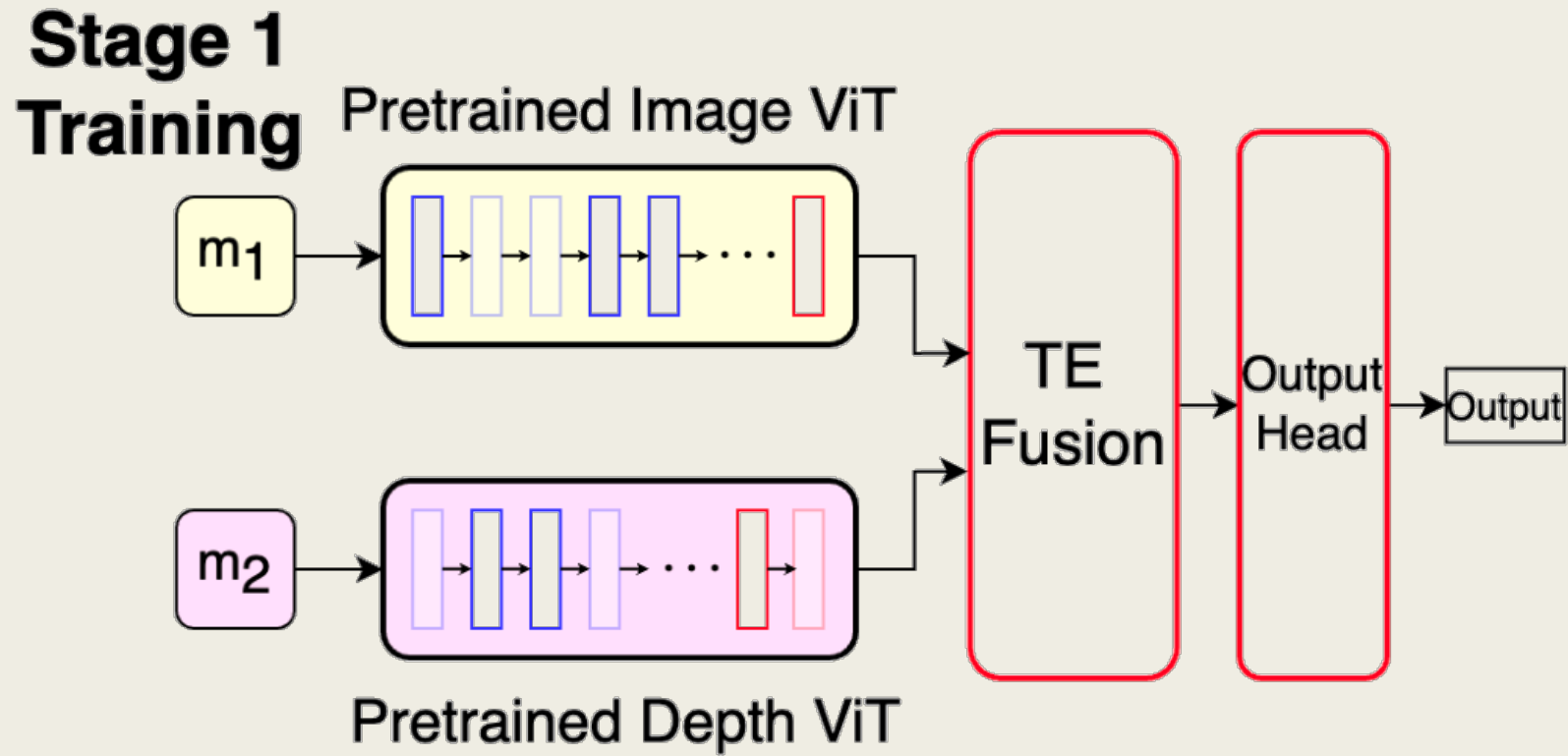


Compute budget: 4 Layers

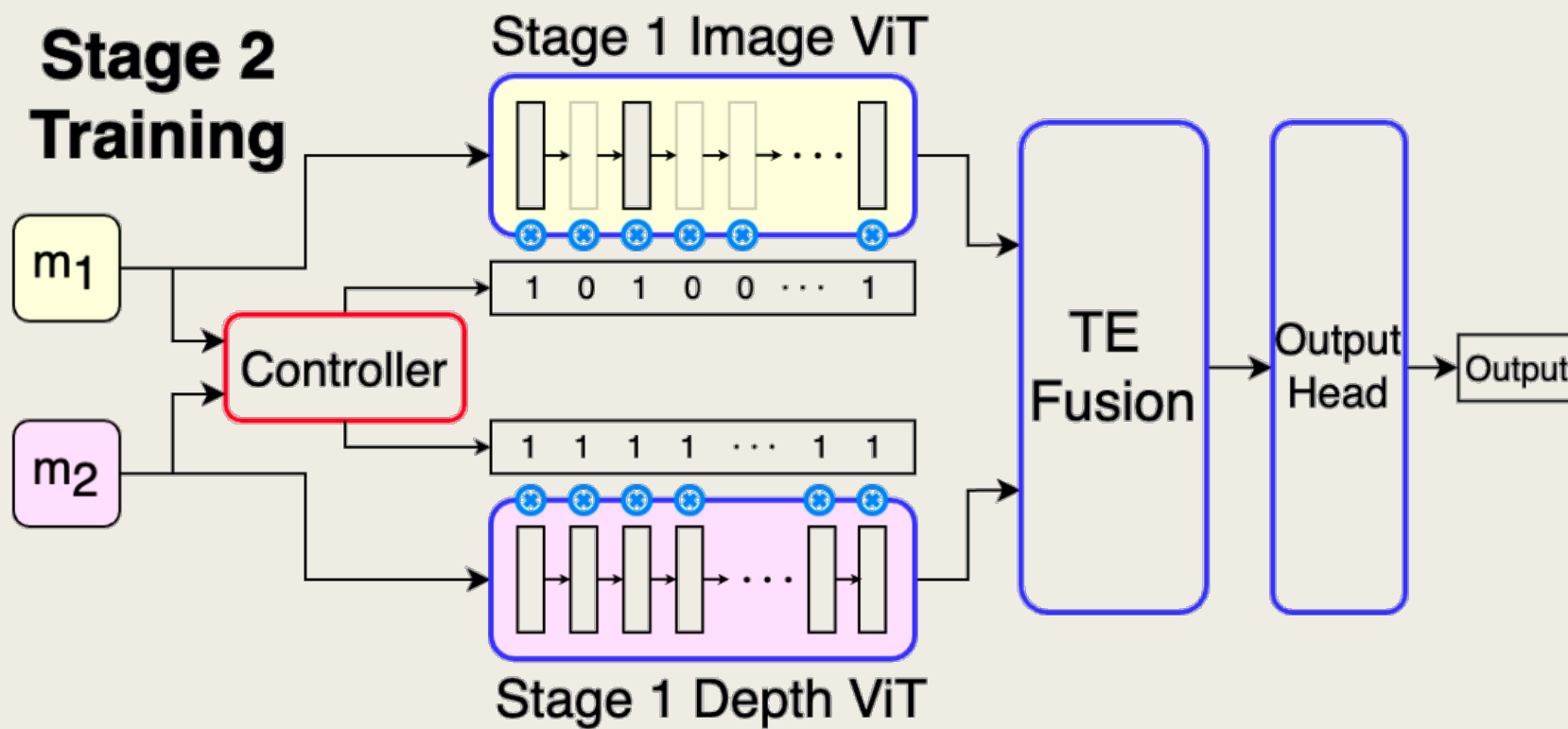


Compute budget: 2 Layers

ADMN Overall Architecture



ADMN Overall Architecture



Main Results – GDTM Localization

Method	GDTM Dataset											
	Gaussian				Lowlight				Blur			
	6	8	12	16	6	8	12	16	6	8	12	16
Upper Bound	29.6				18.8				9.4			
ADMN	51.4	39.0	33.1	<u>30.3</u>	<u>49.5</u>	<u>23.9</u>	18.0	17.3	11.8	11.2	9.8	9.9
ADMN_AE	53.6	38.4	<u>33.5</u>	29.4	35.3	22.8	<u>18.7</u>	<u>17.6</u>	<u>14.0</u>	11.4	<u>9.8</u>	9.6
MNS	<u>39.3</u>	32.1	57.5	73.0	117.5	83.9	117.6	116.8	109.5	114.2	117.7	74.0
Naive Alloc.	112.5	97.6	46.9	31.0	90.3	67.3	27.1	17.7	81.4	50.6	12.5	<u>9.8</u>
Naive Scratch	39.1	<u>36.7</u>	40.4	37.3	117.5	84.6	117.6	117.5	117.8	116.2	117.7	93.6
Image Only	67.5	53.2	53.7	—	56.8	45.6	46.2	—	15.4	<u>11.2</u>	10.3	—
Depth Only	85.0	61.7	58.6	—	72.1	64.9	63.4	—	25.7	22.6	22.3	—

Upper Bound utilizes 24 layers (12 for each modality backbone)

Acknowledgements



National Institutes
of Health